

# MOLECULAR PALAEOLOGY

*by* BRUCE RUNNEGAR

(Twenty-eighth annual address, delivered 15 March 1985)

**ABSTRACT.** Recent developments in molecular biology are beginning to provide new ways of looking at the history of life. At present there are three main potential sources of information: organic molecules extracted from rocks or fossils, the comparative molecular biology of living organisms, and the knowledge that is developing about the role of biopolymers in the construction of skeletons. Each of these fields is reviewed briefly to illustrate how the information that is becoming available may be used in future to serve the common goals of palaeontology and molecular biology.

I LIKE to take the catholic view that palaeontology deals with the history of the biosphere and that palaeontologists should use all available sources of information to understand the evolution of life and its effect on the planet. Viewed in this way the current advances being made in the field of molecular biology are as important to present-day palaeontology as studies of comparative anatomy were to Owen and Cuvier. I appreciate that palaeontologists have just participated in an intellectual and methodological revolution of the first magnitude (the development of global tectonics from the unpopular theory of continental drift) and that it is becoming increasingly difficult to develop interdisciplinary expertise, but the results that are now appearing suggest that molecular biology will be as important to the whole of biology as an understanding of atomic structure was to the physical sciences.

This does not mean that palaeontologists must adopt a passive role as educated observers of this explosion of knowledge. Instead, palaeontologists have the kinds of skills that are required to develop a general understanding of the experimental results that are flooding the literature at the present time. Most molecular biologists have limited training in the classical disciplines of biology and little appreciation of the nature of the fossil record and the dimensions of geological time. Their remarkable experimental and inductive skills will be strengthened through interactions with scientists having an expert knowledge of the history of life and the large-scale processes and effects of evolution.

Needless to say, I am not the first to advocate this approach. Until his untimely death early in 1984, T. J. M. Schopf was a champion of this cause (Gould 1984). He, more than any other person, attempted to bridge the gulf between palaeontology and molecular biology—a difficult and demanding task. At the time of his death we were beginning (at his suggestion) to try to put together the information from molecular biology that might help understand the early history of the Metazoa. I shall refer to that study below; the point to be made here is that Tom Schopf was convinced that it is vital that palaeontologists begin to formulate evolutionary hypotheses that can be tested by further experiments in molecular biology.

As with palaeontology, I take a catholic view of the field known as molecular palaeontology—a term used for many years for the study of ‘chemical fossils’ and incorporating the subject known as ‘palaebiochemistry’ (Abelson 1956; Sylvester-Bradley 1964; Degens 1967; Eglinton and Calvin 1967; Calvin 1968). When Melvin Calvin delivered the Bennett Lecture at the University of Leicester on the topic ‘Molecular Palaeontology’ in 1968 (Calvin 1968) he dealt only with molecules extracted from rocks. This approach has proved to be of fundamental importance to the oil industry (Brooks 1981) and to studies of the early history of life (J. W. Schopf 1978), but the term ‘molecular palaeontology’ can also be far more embracing.

There are three main areas where molecular biology and palaeontology impinge on each other; these three areas constitute what I prefer to call 'molecular palaeontology'. The first is the traditional field of fossil molecules. The second is the role of biopolymers in the construction of the mineral and carbohydrate skeletons that constitute most fossils. The third is the historical information that may be obtained in a quantitative form from comparisons of the primary structures of the proteins and genomes of living organisms. I deal briefly with each of these aspects below. As you will see, there are at present more questions than answers. The great potential of molecular palaeontology has yet to be realized, and in that sense, molecular palaeontology may be compared with isotopic dating and palaeomagnetism at the end of the 1940s.

The literature of molecular biology is overwhelming (*Biochimica et Biophysica Acta* ran to 46 volumes in 1981 and about 10 new journals in the field appeared in the twelve months to March 1985). Consequently, suitable examples may be selected more-or-less at random. I have therefore chosen—so far as is possible—to use examples based upon work done in Australia or upon Australian materials. This fact alone will demonstrate that the treatment is far from comprehensive.

Before proceeding, however, it should be pointed out that the sister discipline of molecular palaeontology is 'atomic palaeontology'. Atomic palaeontology deals (for example) with information obtained from stable isotopes, with the distribution of major and trace elements in skeletal materials, and, of course, with the presence of anomalous amounts of iridium and other noble metals. The potential of these kinds of data to palaeontology may be illustrated by the following simple example.

The ratio of deuterium (D) to hydrogen in the cellulose of woody tissues is thought to diminish in plants living at progressively higher latitudes (Smith *et al.* 1983). Although the causes of this relationship are not well understood (Lawrence and White 1984), Smith *et al.* have been able to show a correlation between the D/H ratios of Australian and Antarctic coals and the palaeolatitudes of their formation. It may therefore be possible to use this technique to determine palaeolatitudes in areas such as Indonesia where the tectonic history and palaeobiogeography are still not well understood (Audley-Charles 1983; Runnegar 1984a).

## BACKGROUND

The fundamental difference between most biological and geological materials is explained simply at the molecular level. In geological systems nearly every atom is linked covalently to other atoms in all directions; in biological systems the covalent bonds are found only along the backbones of linear polymers and the other bonds are weak (Frauenfelder 1983). This immediately explains the softness and flexibility of most biological materials and its highlights the importance of linear polymers to the origin and evolution of life.

A second important point is that biological materials tend to be less ordered than inorganic crystals. According to Galloway (1984), topology is more important than geometry and a biological structure requires no more order than is necessary for it to work. Thus biological materials exhibit some of the properties of crystals and others of liquids. Crystalline order may be present in only one dimension—the one that is required for the structure to function.

A third important point is that linear polymers contain information of the kind not normally expected in three-dimensional crystal lattices (Dose 1983; Matsuno 1983). The sequence of the subunits (residues) provides historical as well as functional information even in molecules such as collagen in which the nature of particular subunits is relatively unimportant. Thus it is the collection and preservation of information that distinguishes life from inanimate objects, and it is this information that is beginning to tell us so much about the way evolution has occurred.

The four main types of linear polymers found in biological systems are nucleic acids (DNA and RNA), proteins, lipids, and carbohydrates. Sugars are particularly unstable under geological conditions so carbohydrates and nucleic acids (in which the bases are linked by sugar rings) are unlikely to survive fossilization (Calvin 1968). Some proteins and many lipids are far more stable and may be isolated (usually in a modified form) from ancient rocks.

## FOSSIL MOLECULES

In addition to the long list of 'prebiotic' organic molecules that have been discovered recently in interstellar space and carbonaceous chondrites (Brown 1980), a large number of different kinds of organic molecules have now been extracted from terrestrial rocks, principally as a result of the development of computerized gas chromatography-mass spectrometry (Ourisson *et al.* 1979, 1982, 1984; Mackenzie *et al.* 1982). This field is of great importance to the petroleum industry (Brooks 1981) and is far too large for this brief review; I shall therefore mention only some of the interesting results of recent studies to illustrate both the potential and the limits of the data.

About 50 years ago, A. Treibs suggested that a common vanadium petroporphyrin (vanadyl deoxyphylloerythroetioporphyrin, DPEP) is the normal geological product of the magnesium-porphyrin complex of the chlorophyll *a* found in all oxygen-producing photosynthetic cells. The vanadyl derivative had been synthesized several times but its identity with the natural product was not confirmed until 1983 when Ekstrom *et al.* (1983) determined the crystal structure of vanadyl DPEP from an Early Cretaceous oil shale (Toolebuc Formation; Saxby 1983) in western Queensland. Similarly, the structure of a nickel petroporphyrin (abelsonite) from the Eocene Green River Formation of Utah was first determined last year (Storm *et al.* 1984); it also appears to be a chlorophyll derivative but the structure alone does not pinpoint derivation from chlorophyll *a*.

Chlorophyll molecules—which consist of the magnesium-porphyrin complex and a long phytol side chain—are supported within protein baskets in the photosynthetic membranes (Thorner and Markwell 1981). The structure of a bacteriochlorophyll *a*-protein association has been determined by X-ray crystallography to a resolution of 0.28 nm (Matthews *et al.* 1979), and a comparison of this structure with DPEP shows how much has been lost during fossilization. Although the demonstration that a 100-million-year-old vanadyl porphyrin is derived from chlorophyll *a* represents a significant achievement in molecular palaeontology (Ekstrom *et al.* 1983), it is clear that only the most stable kinds of biological molecules are likely to be extractable from rocks and that determination of their structure will be a protracted process.

Like the monoplacophoran *Neopilina* and the coelacanth *Latimeria*, there is an important class of fossil molecules that was extracted from rocks before being discovered in the living biota (Ourisson *et al.* 1979, 1982, 1984). These molecules, now known to be derivatives of components of bacterial membranes, are called hopanoids. They indicate that a significant fraction of all crude oils is of bacterial origin. Hopanoids are also abundant in low-rank coals; for example, Ourisson *et al.* (1984) estimated that each cubic metre of an Australian Palaeocene lignite contains about one kilogram of a particular hopanoid acid. Other unusual hopanoids are found in both the Victorian lignites and crude oils from the nearby offshore Gippsland Basin (Philp and Gilbert 1982). These kinds of observations are being used to study the history of the generation and migration of the economically and strategically important Gippsland Basin oils.

All cells are surrounded by membranes (not to be confused with cell walls) that act as dynamic barriers between the external environment and the cytoplasm (Lodish and Rothman 1979). Cell membranes are composed of three main components, lipids, proteins, and carbohydrates. The lipids and proteins are the major components and are present in approximately equal masses, but the protein molecules are much larger than the lipid molecules so lipids are far more numerous.

Lipids are elongate amphipathic structures. This means that they have a hydrophobic end (soluble in oil) and a hydrophilic end (soluble in water). The hydrophobic ends point towards the centre of the lipid bilayer that forms the membrane and the hydrophilic heads face outwards on either side. In mammalian cells there are two main kinds of lipids—flexible molecules (phospholipids) and rigid ones (cholesterol). The rigid cholesterol molecules act as struts to strengthen the membrane whereas the phospholipids allow the membrane to be flexible and, in places, to have a small radius of curvature.

The hopanoids are now believed to be the bacterial analogues of cholesterol (Ourisson *et al.* 1982, 1984). They also have pronounced amphipathic properties and are similar in shape to cholesterol, but the hydrophilic parts of the two molecules are at opposite ends. A large number of derivatives of



bacterial hopanoids have been recovered from sedimentary rocks in the last few years (Ourisson *et al.* 1979; Mackenzie *et al.* 1982).

The point to be made here is that molecular palaeontology has yielded new insights into the nature of bacterial membranes as well as providing an important tool for studies of oil genesis (Mackenzie *et al.* 1982). Because hopanoids are only soluble in mixtures of polar and non-polar solvents (say chloroform and methanol) they were not discovered in living bacteria until a deliberate search for them was made (Ourisson *et al.* 1984).

### *Palaeobiochemistry*

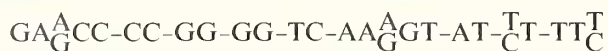
The analysis of molecules obtained from fossils has generally been described as 'palaeobiochemistry' to distinguish such studies from those dealing with molecules dispersed in sedimentary rocks. Palaeobiochemistry also has considerable potential despite somewhat inauspicious beginnings.

Although there have been some partly successful attempts to characterize lipids and nucleic acids from extinct organisms (Niklas *et al.* 1982; Higuchi and Wilson 1984; Higuchi *et al.* 1984), most work in palaeobiochemistry has concentrated on fossil proteins (Abelson 1956; Degens 1967; Armstrong *et al.* 1983). The preservation of objects resembling cell nuclei in silicified cycad wood from the Triassic of New Mexico (Gould 1971) could indicate that some components of nucleic acids may survive in exceptional circumstances, but it is unlikely that much information will be recovered from fossil nucleic acids even if they are found in ancient rocks. On the other hand, even though lipids are more stable than proteins, they contain too little information to be of any real significance except in the ways already explained. Thus fossil proteins offer the best hope for palaeobiochemical work. A variety of techniques have been explored; they include solid-phase radioimmunoassay of collagen from living and extinct vertebrates (Lowenstein 1980)—the technique used to show that the Piltown jaw came from an orangutan (Lowenstein *et al.* 1982), measurements of the extent of racemization of amino acids in skeletal proteins (Kimber and Milnes 1984), and determinations of the concentration of  $\gamma$ -carboxyglutamic acid (Gla) in modern and near-modern bones (King 1978). Each method has its own particular problems and all work best with modern and subfossil materials. Such techniques are therefore likely to be of most use to Quaternary geologists and biologists.

## MOLECULAR FOSSILS

Proteins are linear polymers of amino acids linked by peptide bonds. The 'central dogma' of molecular biology is that the information content of nucleic acids is translated into the amino-acid sequences—the primary structures—of the proteins they specify (Ayala 1978). DNA is transcribed into messenger RNA (mRNA) by enzymes called RNA polymerases and translation of the mature mRNA occurs by an interaction of transfer RNAs (tRNAs) with both the mRNA and disjunct amino acids in ribosomes. In higher organisms (eukaryotes) the genes normally consist of separate coding regions (exons) and non-coding regions (introns). The introns are removed during RNA processing and the ends of the exons are spliced together to make the mature mRNA (Mattaj 1984).

The information potential of nucleic acids and proteins is prodigious. Bodenmüller and Schaller (1981) have demonstrated that an identical 11-amino-acid (head activator) neuropeptide occurs in animals as distant as cnidarians and humans. Although the part of the gene coding this particular neuropeptide has not yet been sequenced, the nature of most of its DNA sequence may be inferred from the genetic code:



(A, adenine; G, guanine; C, cytosine; T, thymine; sites indicated by dashes could be any one of the four nucleotides). A DNA sequence of this form represents one of about  $10^{14}$  possibilities, so even though the neuropeptide is a very small molecule the probability of it having arisen twice by chance is vanishingly small (there are about  $4 \times 10^{13}$  micrometres in the circumference of the Earth). Thus we can be reasonably certain that this short segment of DNA has been passed from generation to



generation in an almost unaltered form since the time of the last common ancestor of cnidarians and vertebrates some 800 million years ago. As such, it represents an extraordinary 'molecular fossil'.

If all polypeptides had changed as little as the head activator neuropeptide in the course of evolution the diversity of life would be low and there would be not much to be learned from a comparison of similar (homologous) proteins of different organisms. However, as the rates of evolution of different kinds of proteins (and their coding sequences) have varied considerably, the comparative biochemistry of homologous proteins is a vast potential source of historical information.

Broadly speaking, proteins may be lumped into three main groups. Many are roughly globular to equidimensional in shape and being water-soluble (hydrophilic) move freely in the cytoplasm. Others are hydrophobic and lie within the lipid bilayers of the cell membranes, and still others are fibrous and serve structural roles (e.g. in muscles and connective tissues). Until recently, most of the published amino-acid sequences were those of the hydrophilic equidimensional proteins because these are more easily extracted and studied. The fibrous proteins tend to have highly repetitive amino-acid sequences that are tedious to determine by traditional methods and the hydrophobic membrane proteins are hard to extract. However, the development of rapid and efficient methods of gene sequencing in the last few years has made available the nucleotide sequences of the genes for a great variety of different kinds of proteins. These nucleotide sequences may be converted into amino-acid sequences using the genetic code.

Just as there are three main groups of proteins, so there are three main kinds of protein structures:  $\alpha$ -helix,  $\beta$ -pleated sheet, and a triple helix typified by the structure of the protein collagen (Richardson 1981; Walton 1981). Many proteins are formed of domains of  $\alpha$  and  $\beta$  structures but others—such as the globins—are dominantly of one type.

The comparative biochemistry of homologous proteins has yielded a large amount of phylogenetic information in the past two decades. Generally speaking, the degree of similarity between homologous proteins of two or more kinds of organisms may be expressed in either qualitative or quantitative terms, but it is the possibility of quantification that has excited the imagination of those interested in the evolution of life. This can be done in a number of ways (for example, by measuring electrophoretic differences), but the most appealing method is a direct comparison of the amount of similarity in the amino-acid sequences of homologous proteins (or nucleic acids). This has led to the development of many different kinds of 'molecular clocks' since the idea was first suggested in 1962 by Pauling and Zuckerkandl (1962; see Wilson *et al.* (1977) for a review). The method has great promise for studies of recent evolutionary events (e.g. the evolution of man; Lowenstein and Zihlman 1984), but from a palaeontologist's point of view an exciting aspect is the potential to look beyond the good fossil record into the vast unknown of the Precambrian. I propose to illustrate this point by discussing briefly some of the molecular and other evidence for the Precambrian history of the Metazoa.

There are three fundamental aspects of the early history of the Metazoa that remain enigmatic. First, are metazoans a monophyletic group descended from a single common multicellular ancestor? Second, are metazoans descended from ciliated protists like *Paramecium*, from other kinds of protists, or indeed, from non-protistan eukaryotes? And third, when did the Metazoa first evolve? Answers to these questions may now be becoming available through the data of molecular biology. For example, the question of the monophyletic versus polyphyletic origin of the Metazoa (Anderson 1982) would seem to be settled by the following evidence.

### *Metazoans are monophyletic*

Collagen is the principal structural protein of metazoan connective tissue and the most abundant protein in higher vertebrates. It has been found in representatives of every metazoan phylum studied and appears to be restricted to the Metazoa (Adams 1978; Towe 1981) although an enzyme required for the post-translational hydroxylation of proline residues seems to have been inherited from the common ancestors of animals and plants (Ashford and Neuberger 1980). Consequently, if the collagens of distantly related metazoan phyla could be shown to be homologous, this would provide powerful support for the idea that all metazoans share a common multicellular ancestor.

There are at least nine different types of vertebrate collagens but the ones of importance for this discussion are those known as the fibrillar collagens (Types I to III). They occur in a variety of tissues including skin, liver, bone, and cartilage (Bornstein and Sage 1980).

Fibrillar collagens are composed of long triple helices formed of identical or homologous polypeptides having the repetitive amino-acid sequence  $(G-X-Y)_n$ , where G is glycine and X and Y are usually proline, alanine, or a charged residue. A post-translational conversion of many of the proline residues to hydroxyproline is required to stabilize the secondary and tertiary structures of the molecules, and it is significant in another context that this post-translational modification requires molecular oxygen (Towe 1970, 1981).

The collagen triple helices, and their short terminal non-helical segments, are overlapped to form fibrils by a distance  $nD$ , where  $D = 234$  amino acid residues (Woodhead-Galloway 1980). Hydrophobic and electrostatic interactions between adjacent triple helices are maximized when the molecules are staggered in this way, and the tensile strength of the whole fibril is provided by the development of covalent bonds between adjacent triple helices.

When collagen fibrils are positively stained with heavy metals for electron microscopy the stain accumulates at the sites of the charged residues and a distinctive banding pattern results (Woodhead-Galloway 1980). Because fibrils are composed of triple helices overlapped by 234 residues the banding pattern has a repeat distance (67 nm) that is equal to the average distance between adjacent residues (0.286 nm) multiplied by 234. However, if the molecules are separated from each other and then recombined so that they lie in register side by side, positive staining of the (SLS) aggregate reveals a banding pattern which is essentially a map of the distribution of charged residues in the molecules (text-fig. 1).

Because the amino-acid sequences of several different kinds of vertebrate fibrillar collagens have been determined either by conventional methods or by gene sequencing (Runnegar, in press *a*), computer-drawn plots of the charged residues may be used to simulate the patterns observed in the SLS aggregates (text-fig. 1). It is, therefore, possible to make a direct visual comparison between the amino-acid sequences of different collagens using either photographs of positively stained SLS aggregates or computer-drawn maps of the amino-acid sequences.

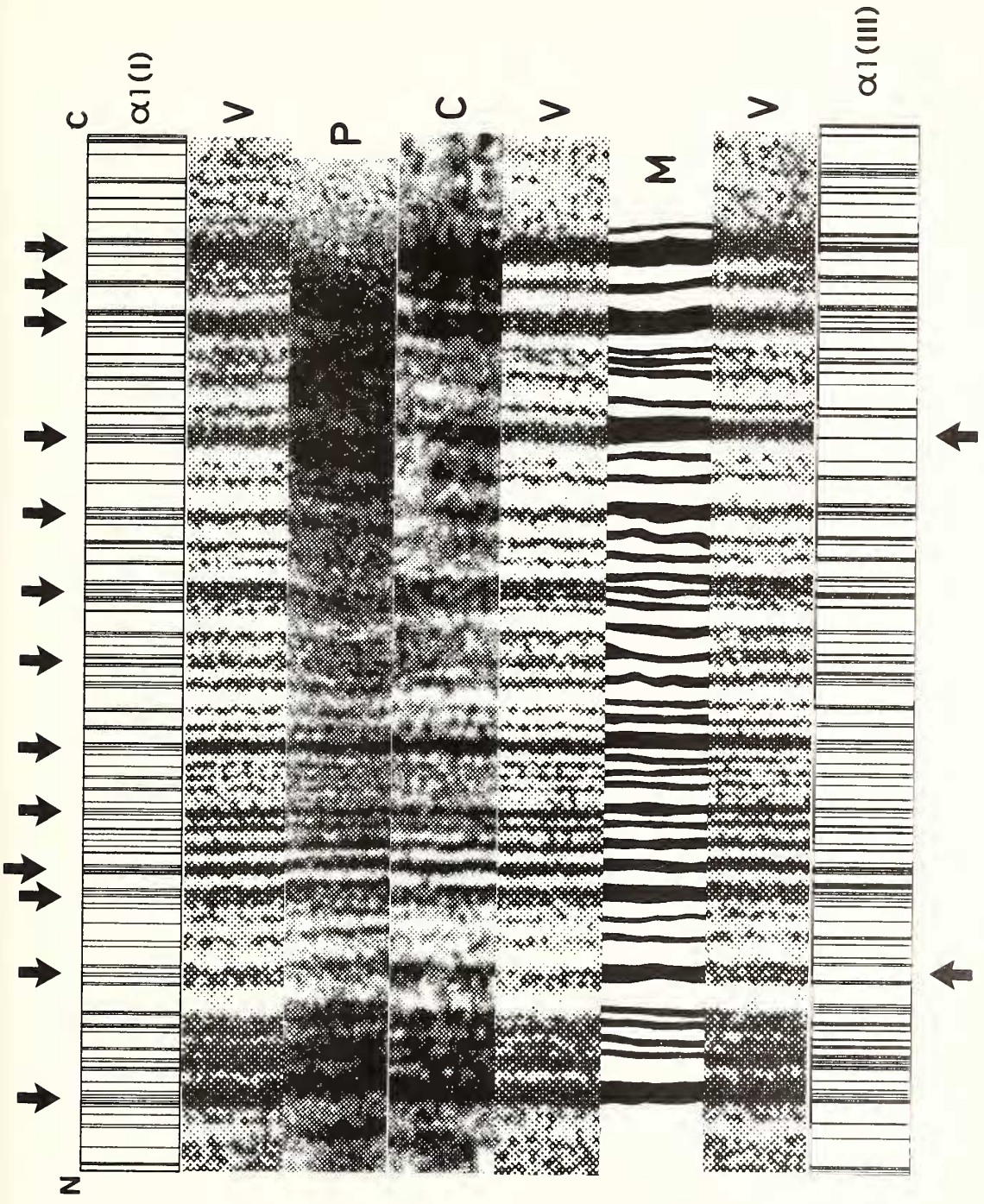
It has been known for some time that the SLS banding patterns of fibrillar collagens from the mesogloea of the cnidarian *Actinia equina*, the body wall of the parasitic platyhelminth *Fasciola hepatica*, and various vertebrates are almost identical (Nordwig and Hayduk 1969). It has recently been shown that collagen from the byssus of the mollusc *Mytilus edulis* has an SLS banding pattern like that of vertebrate Type I collagen (DeVore *et al.* 1984). These similarities are obvious in text-fig. 1. Furthermore, the invertebrate collagens are more similar to vertebrate Type I collagens than they are to vertebrate Type III collagens. Thus there is a greater similarity in the collagens of distant phyla of the Metazoa than there is between collagen molecules that may be covalently cross-linked in a single tissue (Henkel and Glanville 1982).

Because the function of collagen molecules is to resist tension, they are designed and act like ropes. The repetitive nature of their amino-acid sequences results from fact that glycine is the only residue

---

TEXT-FIG. 1. Distribution of charged residues in the telopeptides of homologous vertebrate and invertebrate collagens. The top and bottom bars are computer-drawn representations of the rat + calf  $\alpha(1)$ I and calf skin  $\alpha(1)$ III sequences (references in Runnegar, in press *a*) and the other bars are enlarged copies of published electron micrographs of positively stained SLS aggregates, as follows: V, vertebrate Type I collagen (after Bentz *et al.* 1978, fig. 3, republished with permission); P, C, platyhelminth body-wall collagen and cnidarian mesogloea collagen (both after Nordwig and Hayduk 1969, pl. 7, republished with permission); M, mollusc byssus collagen (traced from photographic enlargement of fig. 3A of DeVore *et al.* 1984); the vertebrate collagen is repeated for clarity. Arrows at top point to clusters of charged residues that are conserved in all molecules; the arrows at the bottom point to clusters of charged residues that are present in vertebrate Type I collagens and the invertebrate collagens but not in vertebrate Type III collagens.







small enough to fit into the axis of the triple helix. Apart from this constraint and the need for certain proportions of hydroxyproline and charged residues, there would appear to be no particular reason—other than a historical one—for the charged residues to be distributed in the way that they are. Collagen genes appear to have been constructed by the serial repetition of an original 54-nucleotide module (Yamada *et al.* 1980; Runnegar, in press *a*), so the disorder evident in the distribution of charged residues represents a unique subsequent development. It is therefore clear that living vertebrates, cnidarians, platyhelminths, and molluscs have inherited homologous collagen genes, presumably from a remote common metazoan ancestor. This is excellent evidence that the metazoans represent a monophyletic clade.

There is, however, one possible flaw in this logic. There is now ample evidence that copies of DNA sequences may be transferred from one genome to another. Such transfers are common within the cells of single organisms; for example, an ATPase gene has migrated from the mitochondria to the nucleus of baker's yeast, the mitochondrial and nuclear genomes of rat liver cells have a common sequence about 3000 nucleotides (nt) in length, and there are extensive homologies between the mitochondrial and chloroplast DNAs in some higher plants (Hadler *et al.* 1983; Stern and Palmer 1984). Gene transfer may also occur between organisms by means of viruses or bacterial plasmids.

It is, therefore, at least conceivable that the present distribution of fibrillar collagen genes is due in part to lateral gene transfer. This explanation is improbable for two reasons. First, collagen is a vital material for all metazoans so any lateral gene transfer would have had to have coincided with the development of multicellularity; this scenario therefore requires the improbable synchronous occurrence of two events. And second, fibrillar collagen genes are large and complex. For example, the chicken  $\alpha(2)I$  gene contains 49 exons and non-coding regions that are about  $3.4 \times 10^3$  nt in total length (Tate *et al.* 1983). It seems unlikely that such a complex gene could be transferred intact to another genome. When the problems of gene expression after transfer are considered as well, the possibility of lateral transfer of collagen genes becomes remote. This may not be true for smaller and simpler genes.

#### *Graptolite collagen genes*

All of the collagens discussed so far were obtained from living animals. But can anything be learned about the collagens of long-extinct organisms? Surprisingly, the answer appears to be yes, because it is becoming apparent that properties of genes are reflected in structures that may be observed in well-preserved fossils.

In collagen genes, most of the exons that encode the triple helical part of the protein molecule are small integral multiples of 54 nt in length (Tate *et al.* 1983). The others are either 45 (54–9) or 99 (54 × 2–9) nt in size and are believed to have been shortened from an original length of 54n nt by the removal of a segment coding for one G–X–Y amino-acid triplet. This explains why collagen genes are thought to have evolved by the tandem repetition of 54-nt modules (Yamada *et al.* 1980; Runnegar, in press *a*).

Only part of a single *Drosophila* collagen gene has so far been sequenced (Monson *et al.* 1982), but at least one of the two exons present in the fragment appears to have been 702 (54 × 13) nt in length prior to the deletion of a few nucleotides (Runnegar, in press *a*). When this fact is coupled with repetitions discovered by McLachlan (1976) in the amino-acid sequence of a vertebrate Type I collagen, it seems likely that collagen genes were also constructed by the successive duplication of 702-nt secondary modules (McLachlan 1976; Runnegar, in press *a*). This explains the origin of the D-period in collagen fibrils (702 nt = 234 amino acids).

Because the collagen triple helices are not integral multiples of 234 amino acids in length, there are spaces (holes) between the ends of the triple helices (Woodhead-Galloway 1980). These holes become filled with heavy metal stains with the fibrils are negatively stained for electron microscopy, and they become filled with apatite when collagen is mineralized (Berthet-Colominas *et al.* 1979). Thus negatively stained collagen fibrils exhibit an alternation of light and dark crossbands under the electron microscope and the dark bands correspond to the positions of 'hole zones' (Woodhead-

Galloway 1980). As a result of the geometry of packing, each pair of light and dark bands is 67 nm (234 amino acids) in length.

Freeze-fracture replicas of unstained collagen fibrils display similar crossbands because the 'hole zones' are less voluminous than the intervening regions (Leonardi *et al.* 1983). Thus the fundamental construction of collagen genes may be determined from measurements of the morphology of essentially untreated collagen fibrils. It is merely necessary to know the inter-residue spacing of a synthetic polymer of the collagen type (0.285 nm in poly L-prolyl-glycyl-L-proline; Traub and Yonath 1966) and the period of the cross band (67 nm) to determine—with the advantage of hindsight—that collagen genes are constructed from 702 and/or 54-nt modules ( $67 \text{ nm}/0.285 \text{ nm} = 235 \times 3 \approx 702 = 54 \times 13$ ; a more precise value for the inter-residue distance in unstretched tendon collagen (0.2866 nm; Fraser *et al.* 1979) gives a better result: 233.8 residues, 701.3 nt).

How does all this relate to palaeontology? It turns out that the same kinds of deductions can be made from molecular structures observed in the graptolite periderm. Towe and Urbanek (1972), Urbanek and Towe (1974, 1975), and Crowther and Rickards (1977) have illustrated banded fibrils in the cortical layers of Late Ordovician specimens of *Dictyonema*. These fibrils have been interpreted as the remains of original collagen, both on the basis of their morphology (Towe and Urbanek 1972) and on the spacing of distinctive crossbands (Crowther and Rickards 1977). More recently, Armstrong *et al.* (1984) have shown that the extra-cellular tubes of living pterobranchs are collagenous in composition, thus supporting the earlier interpretations of the structures found in graptolite skeletons and also the hypothesis that the graptolites are closely related to the pterobranchs.

Crowther and Rickards were not particularly interested in the exact value of the periodicity in the crossbands of the fibrils of *Dictyonema* since a value of about 70 nm was sufficient to establish the collagenous nature of the material. However, measurements made from their published photographs suggest that the repeat distance lies between 65 and 70 nm. It is, therefore, likely that the triple helical molecules of the cortical collagen of *Dictyonema* were staggered by 67 nm and that *Dictyonema* collagen genes were constructed from 54 nt modules. Thus graptolite collagen appears to have been homologous to the collagens illustrated in text-fig. 1.

The crossbands of graptolite collagens are visible in ultra thin sections and therefore cannot be merely a topographical feature (Towe and Urbanek 1972, fig. 4). The alternation of light and dark bands is reminiscent of the pattern seen in negatively stained fibrils and so it is possible that the electron-dense regions represent hole zones that have been partially mineralized during preservation and diagenesis. If introduction of mineral into these regions has mimicked the effects of negative staining and biomineralization, it may be possible to find traces of collagens in other invertebrate fossils and hence to use the crossbanding patterns to learn something about the nature of their collagen genes.

#### *Dating the origin of the Metazoa*

Differences in the sequences of homologous proteins and nucleic acids from different animal phyla may—at least in theory—be used to date the times of origin of the various animal phyla. However, the amount of useful information so far available is limited and the only question worth addressing at present is whether the animal phyla have a short or long Precambrian history (Sepkoski 1978; Runnegar 1982a).

The method involves a quantitative comparison of the residues of homologous biopolymers (usually given as percent difference); a correction for substitutions that have reverted to the original condition and thus appear unchanged whereas they have changed twice, and a calibration of the rate of evolution based upon an event that can be identified (and isotopically dated) in the fossil record. Thus, the method is not simple and it involves a number of factors that are difficult to determine.

There is an additional complication in that it is necessary to deal with two kinds of divergence—the divergence of lineages and the divergence of genes. Duplicate copies of genes within organisms begin

to evolve separately once duplication has occurred. Some copies become so modified as to be unserviceable and may remain in the genome as non-functional pseudogenes. For example, the human genome contains two closely related embryonic  $\alpha$ -like globin genes, only one of which encodes a functional polypeptide (Proudfoot *et al.* 1982). The protein-coding regions of the two genes differ in only three nucleotides but one of these mutations has produced a termination codon in the non-functional gene. The protein-coding sequence can therefore not be translated into a globin molecule and it resides in the genome as an unexpressed pseudogene.

It is clear from this simple example that genes have their own histories of origin, evolution, and extinction. Perhaps all modern genes are copies that have been re-copied many times, not only from generation to generation, but from place to place within evolving genomes. The extra copies appear to serve three main functions: they enable their products to be manufactured quickly; they represent a safeguard against failure through mutation; and they provide scope for experimentation (one copy can evolve while another continues to manufacture a vital product). The possibility that unexpressed pseudogenes may return to a functional role after a period of evolution remains little more than an idea but it exemplifies the way molecular biology is changing the way we look at evolutionary mechanisms.

As well as giving an indication of the age of the Metazoa, the respiratory pigments known as globins illustrate the basic principles of molecular evolution. At present, globins are probably the best examples available because they can be shown to be homologous, they have evolved at an intermediate rate, and a large number of complete amino-acid sequences are available.

The protein part of a globin molecule resembles a framework constructed from unequal lengths of pipe joined by U-pieces. The pipe-like parts are lengths of  $\alpha$ -helix, and, as most of the molecule has this kind of structure, there are limited constraints on the nature of many of the 140 or so amino-acid residues. The 'works' of a globin—the part that reversibly binds oxygen—is an iron-porphyrin complex not very different from the magnesium-porphyrin complex of the chlorophylls. It lies within the protein frame and is held in place by the side chains of greatly conserved amino-acid residues (Dickerson and Geis 1983).

Because the segments of  $\alpha$ -helix are unequal in length the tertiary structure of globin molecules is quite irregular. The same irregular structure is present in globins from vertebrates, an annelid, an insect, and the root nodules of a legume (Lesk and Chothia 1980). When these features are coupled with the common characters found in globin genes and their amino-acid sequences, there can be little doubt that all globins are homologous (Runnegar 1984b).

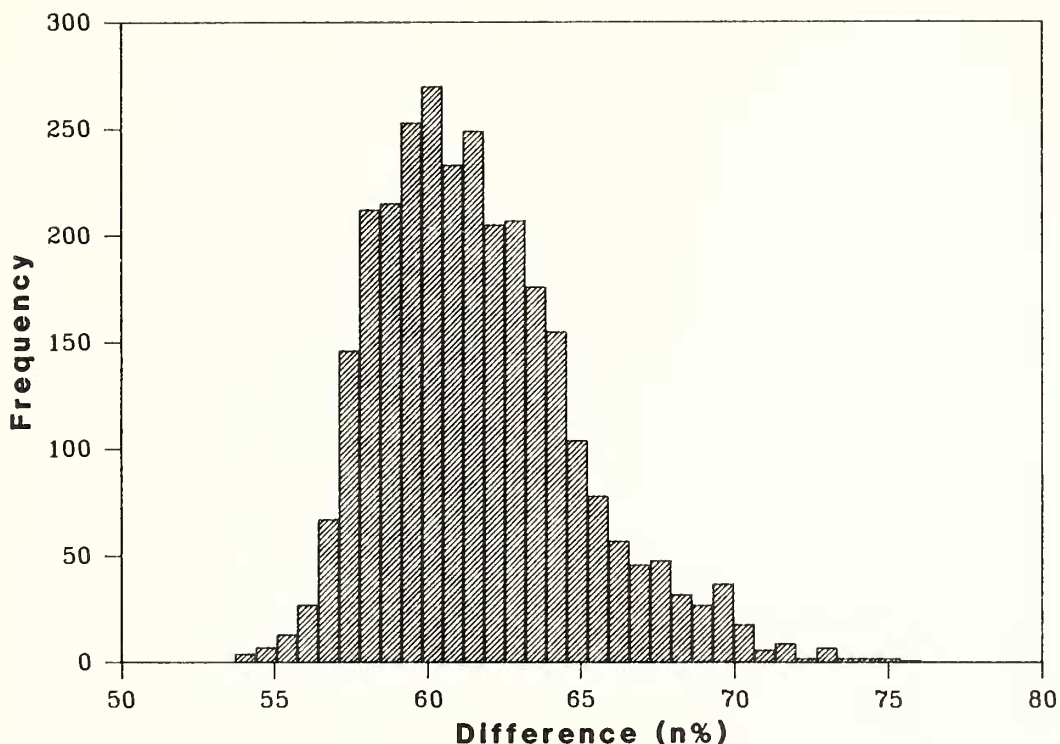
Globin molecules found in muscle cells (myoglobins, Mb) are monomeric, but the globins that circulate in body fluids (haemoglobins, Hb) are generally either intracellular small polymers (commonly four subunits) or large extracellular polymers of as many as 186 subunits (Messerschmidt *et al.* 1983). In all living vertebrates except jawless fish the main component of haemoglobin is a tetramer formed of two pairs of distantly related globin monomer called the  $\alpha$  and  $\beta$  chains.

The primordial  $\alpha$  and  $\beta$  haemoglobin genes were produced by a gene duplication that post-dated the evolution of the jawless fish during or prior to the late Cambrian. The duplication occurred in the lineage leading to all other vertebrates, including sharks. It is, therefore, possible to date this gene duplication event to about 450 million years ago (middle-late Ordovician) and to use this date to calibrate the rate of evolution of globin molecules.

Once the gene duplication had occurred and  $\alpha$  and  $\beta$  genes began to evolve independently. Each living organism possessing these genes has had the same amount of time for evolution to occur, so the amount of difference between the amino-acid sequences of any  $\alpha$  and  $\beta$  haemoglobin should be exactly the same if the molecular clock has any meaning. Any departures from equivalent amounts of difference may then be attributed to variations in the rate of change, or perhaps, to a greater tendency to revert to the original condition.

As might be expected, a comparison of a large number of  $\alpha$  and  $\beta$  haemoglobin sequences reveals that some pairs are more alike than others. However, a histogram of the frequency of the values obtained from pairwise sequence comparisons displays all of the characteristics of a normal



$\alpha/\beta$  haemoglobin differences. All comparisons.

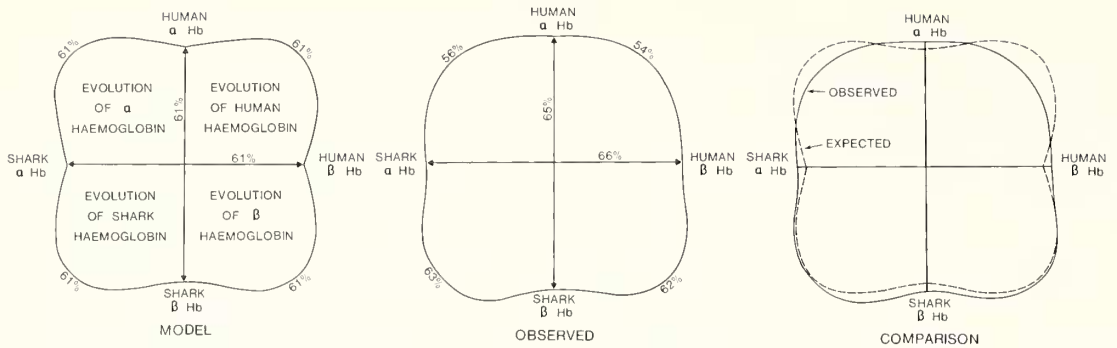
TEXT-FIG. 2. Histogram of the observed percentage differences obtained from the pairwise comparison of the amino-acid sequences of vertebrate  $\alpha$  and  $\beta$  haemoglobins ( $N = 2915$ ,  $\bar{x} = 61.6$ ,  $S = 3.2$ ). The distribution is statistically insignificant different from normal except for a small amount of skewness.

distribution except for a small but statistically significant amount of skewness (text-fig. 2). This distribution illustrates the point that molecular clocks are 'sloppy' and that results obtained from comparisons of only a few sequences are likely to be misleading. On the other hand, the average difference obtained from the 2915 sequence comparisons used for text-fig. 2 is 61.6%; this value is not very different from a mean value of 61.05% obtained previously from only eighty sequence comparisons (Runnegar 1982a).

The problems of correcting for superimposed mutations (ones that have resulted in the restoration of the original condition), and for the fact that certain amino acids are more likely to be replaceable than others, is beyond the scope of this brief review (see Golding 1983 for a recent discussion). There is, however, a need for some kind of correction for superimposed mutations and the simple method described in Runnegar (1982a) will be used here.

It is, of course, often claimed that the rates of evolution of different proteins have varied at different periods of time. For example, the rates of evolution of higher primate globins are thought to have been exceptionally slow because a molecular clock date for the origin of man is much too young. On the other hand, some authors have suggested that proteins evolve quickly when they first appear and that the rate of evolution slows down subsequently. Other argue for alternations of fast and slow rates (Goodman 1981).

It seems possible that rates of evolution obtained from proteins and nucleic acids may be unreliable when based upon small samples or closely related molecules. With larger samples and/or longer



TEXT-FIG. 3. Comparison of the observed and expected differences between the amino-acid sequences of shark and human  $\alpha$  and  $\beta$  haemoglobins. See text for further explanation. The relatively small difference between the  $\alpha$  and  $\beta$  haemoglobins of humans is unusual (text-figs. 2 and 4).

periods of time, the molecular clock seems to work, at least in an approximate fashion. For example, the potential of the globin clock may be illustrated in the following simple way:

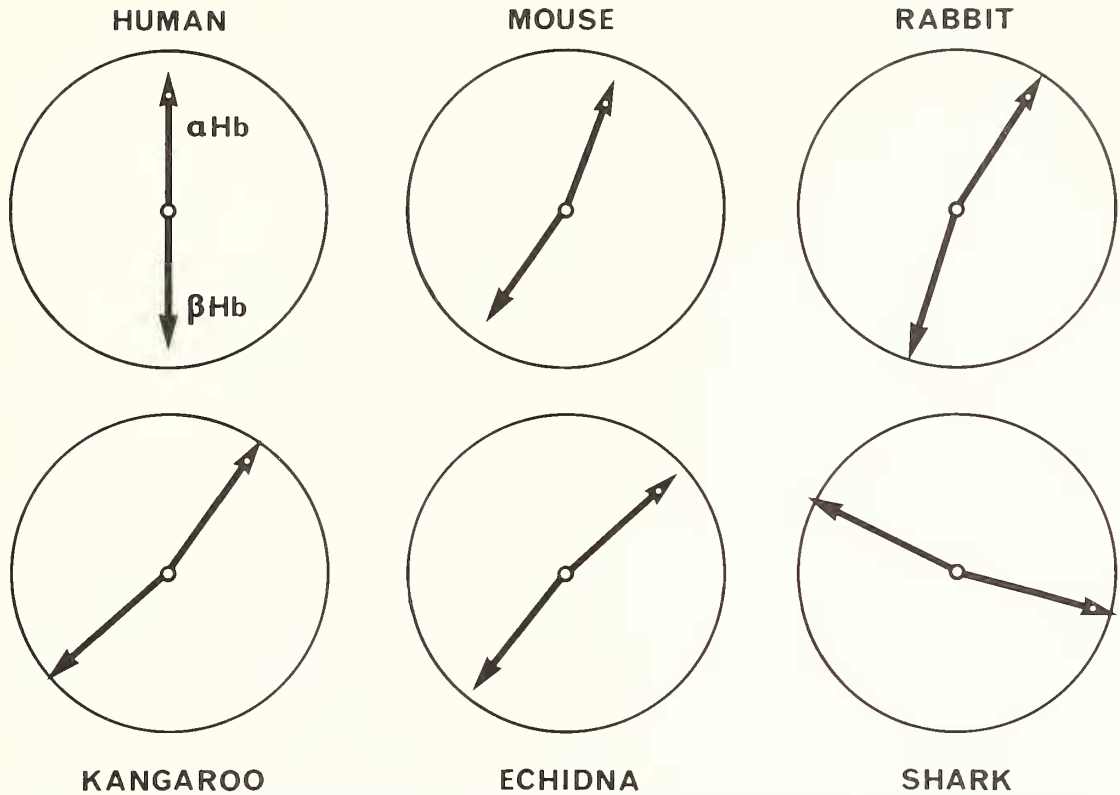
1. The last common ancestor of humans and modern sharks was a Late Ordovician or Silurian fish that had inherited the recently acquired duplicate genes for the  $\alpha$  and  $\beta$  chains. Therefore, the  $\alpha$  and  $\beta$  globins of the shark have been isolated from their human counterparts for almost as long as the duplicate genes have been evolving independently. Consequently, it is not surprising that there is almost the same amount of difference between the  $\alpha$  globins of sharks and humans as there is between their  $\beta$  globins, and that both figures are close to the average difference (61%) between the  $\alpha$  and  $\beta$  globins of living vertebrates (text-fig. 2). It is, therefore, possible to model the expected results and to compare observed with expected values (text-fig. 3). The close fit supports the idea that shark globins have evolved at much the same rate as those in the vertebrate lineage leading to man, despite the fact that the physiological requirements of sharks and mammals are quite different.

2. The evolution of the  $\alpha$  and  $\beta$  globins may be illustrated diagrammatically in clock form (text-fig. 4). The diameter of the face of the clock may be used to represent the average percentage sequence difference between the  $\alpha$  and  $\beta$  globins of living vertebrates and the length of the hands can represent the observed percentage sequence difference in each particular case. If hands representing the human  $\alpha$  and  $\beta$  globins are placed at 12 and 6 o'clock, and the comparative sequence differences in other globins are shown as the clockwise distance away from the human position, three things are obvious. First, the sequence difference between  $\alpha$  and  $\beta$  globins is similar in all six animals; second, both kinds of globin depart by a roughly equal amount from their human counterparts; and third, the amount of sequence difference corresponds well with the evolutionary distance from humans.

3. An earlier duplication of a vertebrate myoglobin gene led to the evolution of the vertebrate haemoglobins from one of the duplicates (Dickerson and Geis 1983). It is, therefore, to be expected that the average sequence difference between  $\alpha$  globins and vertebrate myoglobins will be equal to the average difference between  $\beta$  globins and the myoglobins. This turns out to be the case; the values are 74.1% and 72.9% respectively when compared over the same number of residues (text-fig. 5).

There are now a number of invertebrate globin sequences available (Runnegar 1984b). Three-way comparisons of annelid, mollusc, and vertebrate sequences show that between-phylum differences (about 80%) are greater than those found within the vertebrates (text-fig. 5).

The results shown in text-fig. 5 are based upon more than a million amino-acid residue comparisons. If the mean values obtained from each set of comparisons are corrected for superimposed substitutions it is possible to use the corrected values to estimate the date of the gene duplication that produced the ancestral vertebrate haemoglobin gene from a pre-existing myoglobin



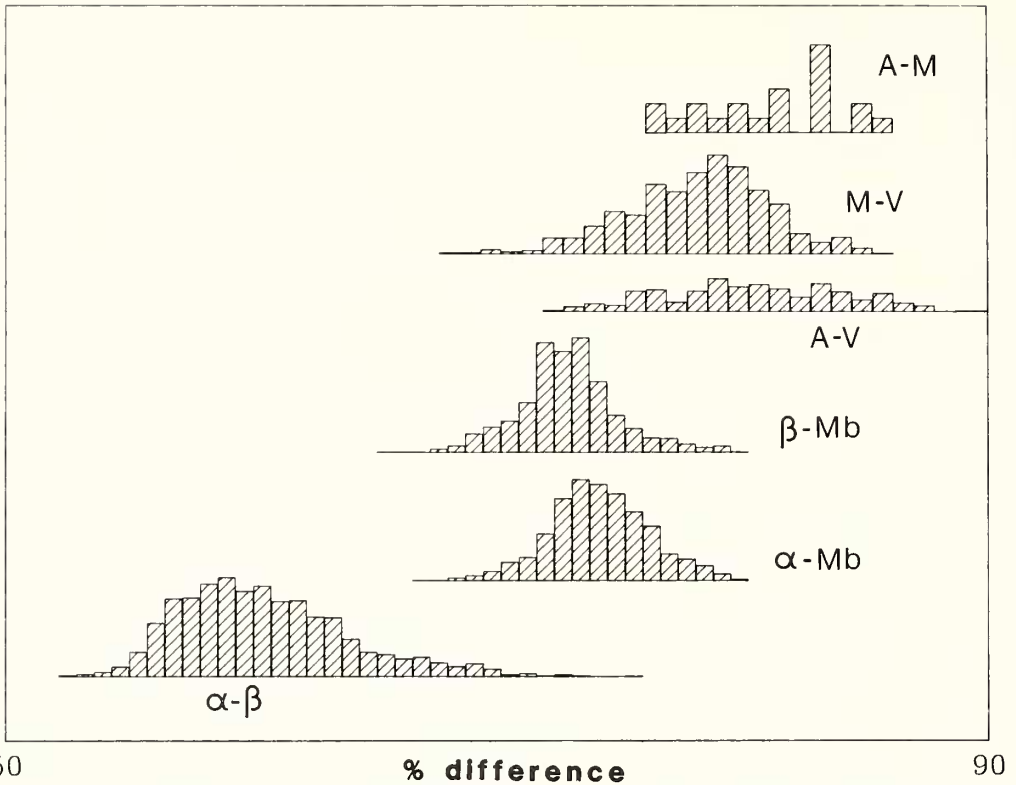
TEXT-FIG. 4. Molecular evolution of selected vertebrate  $\alpha$  and  $\beta$  haemoglobins shown in clock form. See text for further explanation. From Runnegar (1982c, fig. 5), republished with permission.

gene, and also to derive an approximate minimum date for the initial radiation of the animal phyla (text-fig. 6). The logic is as follows.

The rate of evolution of the globins is calibrated by the gene duplication that produced the ancestral  $\alpha$  and  $\beta$  chains in the Ordovician as described above and in Runnegar (1982a). As the vertebrates are monophyletic and did not originate until the late Cambrian (Briggs and Fortey 1982), their early Cambrian and Precambrian history comprises a single species-lineage. The divergence-times obtained from  $\alpha$ -Mb and  $\beta$ -Mb comparisons therefore date this event within that lineage. In other words, 'vertebrate' haemoglobin first appeared in a direct ancestor of the Vertebrata during the Ediacarian (text-fig. 6). This ties in fairly well with the idea that respiratory transport pigments—as distinct from muscle storage pigments—evolved during the Ediacarian in response to increasing amounts of free oxygen in the atmosphere and hydrosphere (Runnegar 1982b, c).

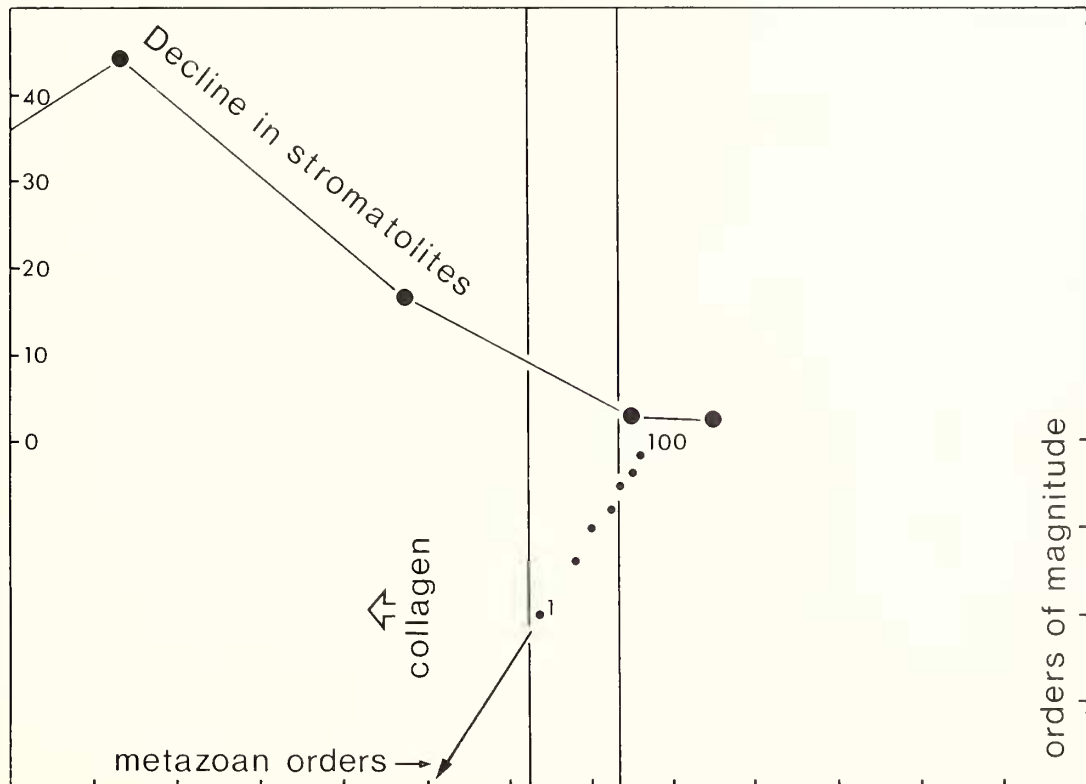
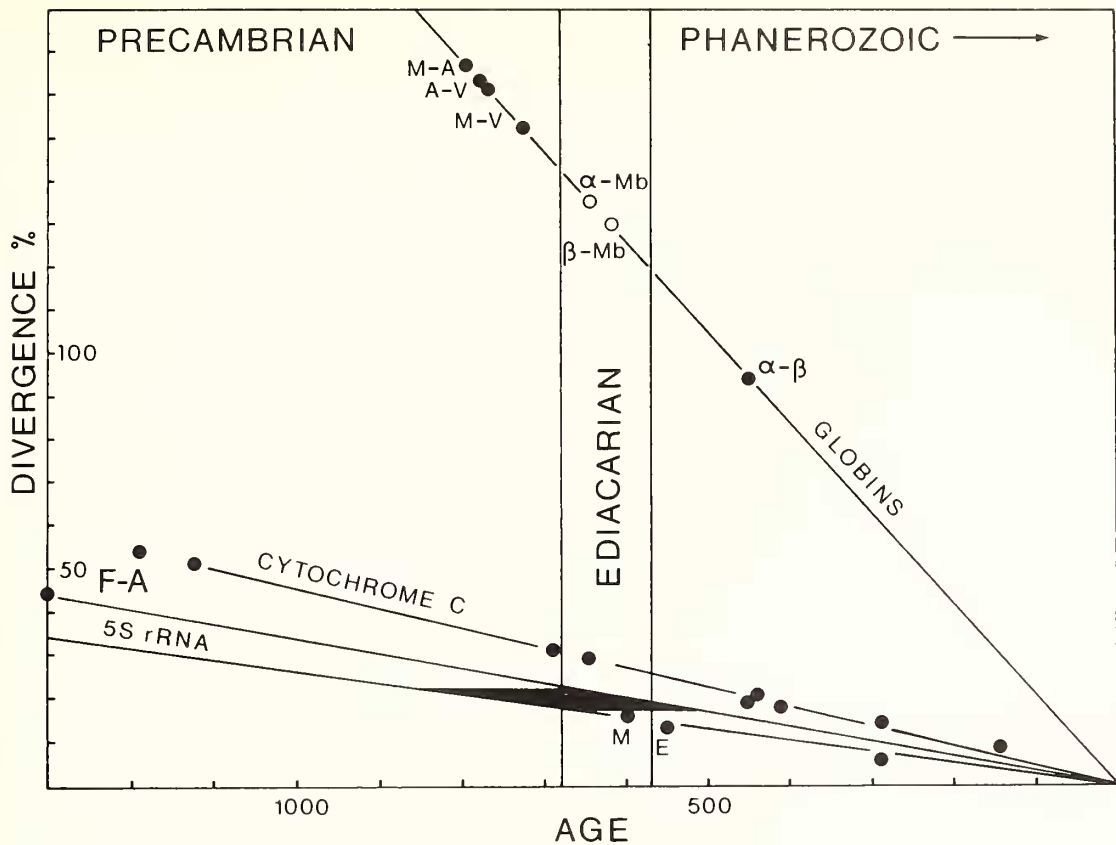
A parallel development probably took place within the lineages leading to other animals phyla (e.g. the Mollusca), but there is at present too little information for a similar analysis of invertebrate globins. However, if all globins are monophyletic, between-phylum comparisons may be used to obtain an approximate minimum date for the initial radiation of the animal phyla. The value of about 800 million years ago shown in text-fig. 6 is likely to be an underestimate for two reasons: first, only homologous residues were used for the sequence comparisons and all unmatched segments of the molecules were excluded; and second, there is an upper limit to change which may be being approached in such different amino-acid sequences. Thus the limited evidence available from the globin clock points to a Precambrian history of the Metazoa of the order of 200–400 million years (text-fig. 6; Runnegar 1982a; see Gingerich (1984) for a different interpretation).





TEXT-FIG. 5. Histograms of the observed percentage differences obtained from the pairwise comparison of the amino-acid sequences of vertebrate and invertebrate globins (plotted at different vertical scales). A-M, annelid/mollusc,  $N = 21$ ,  $\bar{x} = 81.2$ ,  $S = 2.8$ ; M-V, mollusc/vertebrate,  $N = 987$ ,  $\bar{x} = 78.2$ ,  $S = 2.9$ ; A-V, annelid/vertebrate,  $N = 423$ ,  $\bar{x} = 80.4$ ,  $S = 3.6$ ;  $\beta$ -Mb, vertebrate  $\beta$  haemoglobin/vertebrate myoglobin,  $N = 1749$ ,  $\bar{x} = 72.9$ ,  $S = 2.0$ ;  $\alpha$ -Mb, vertebrate  $\alpha$  haemoglobin/vertebrate myoglobin  $N = 1815$ ,  $\bar{x} = 74.1$ ,  $S = 2.0$ ;  $\alpha$ - $\beta$ , vertebrate  $\alpha$  haemoglobin/vertebrate  $\beta$  haemoglobin as in text-fig. 2.

TEXT-FIG. 6. A, molecular estimates of the time of origin of the metazoan phyla based upon the data given in text-fig. 5 (globins) and on differences in the amino-acid sequences of cytochrome *c* and the nucleic-acid sequences of 5S rRNAs. The rates of evolution are based upon dated events within the Phanerozoic such as the origin of the genes for  $\alpha$  and  $\beta$  haemoglobins, the origin of the echinoderm classes (E) and times of divergence of various vertebrate groups (fish/mammals, birds/mammals, etc.). Because 5S rRNA molecules have evolved slowly it is difficult to calibrate their rate of evolution from the information currently available. A distant calibration point may be provided by the average difference between fungal and animal 5S sequences (F-A), on the assumption that these two kingdoms last shared a common ancestor about 1300 million years ago (a somewhat younger date is given by the cytochrome *c* data). A slower rate of 5S rRNA evolution is indicated by comparisons between molluscan and echinoderm classes (M, E). The solid field indicates the limits of points derived from between-phylum comparisons; the spots above this field were obtained from between-phylum comparisons of cytochrome *c* sequences. B, observed decline in abundance/diversity of late Proterozoic and Cambrian stromatolites (Walter and Heys, in press), probable minimum time of origin of vertebrate collagen genes (Runnegar, in press *a*), and a backwards extrapolation of Sepkoski's (1978) estimate of the number of metazoan orders in the Ediacarian and Cambrian. The extrapolation is based upon the premiss that between 1 and 10% of metazoan orders existing at the time were fossilized.



The globin data are supported in a limited way by data from amino-acid sequences of the protein cytochrome *c* and the nucleotide sequences of small ribosomal RNA molecules (5S rRNAs; text-fig. 6). In the case of the 5S rRNA sequences, the rate of evolution may be calibrated in two ways, either by using limited data from different classes of molluscs and echinoderms (M. E., text-fig. 6) or by using data from fungi which appear to have diverged from the lineage leading to the animals some 1200–1300 million years ago. Because of the very limited number of sequences available the estimates of the times of divergence of the animal phyla shown in text-fig. 6 should not be taken too seriously; they are presented here more to illustrate the technique than to provide an answer to the problem.

Another way of looking at this problem is to assume that the Ediacarian and Cambrian fossil record is likely to contain between one and 10% of the higher taxa that existed at the time. If so, it may be possible to extrapolate Sepkoski's (1978) curve of the diversity of Ediacarian–Cambrian marine orders backwards into the Precambrian through one or two orders of magnitude (text-fig. 6). The answer given by this (admittedly dubious) extrapolation is comparable to that obtained from the molecular evidence. Such a date is also partly supported by new estimates of the diversity and abundance of Precambrian and Cambrian stromatolites (Walter and Heys, in press; text-fig. 6); the substantial decline in both diversity and abundance that began between about one billion and 800 million years ago is attributed to grazing by newly evolved metazoans.

Finally, some evidence of the time of origin of the metazoan phyla may also be obtained from collagen molecules. The sequence data so far available are not ideal as they mostly come from animals (birds and mammals) that have diverged relatively recently. Nevertheless, there is some indication that the genes for Type I and Type III collagens diverged about 800–1000 million years ago (Bernard *et al.* 1983; Runnegar, in press *a*). This event may well have occurred early in the history of the Metazoa, but until more invertebrate or lower vertebrate sequences become available it will be difficult to test this hypothesis and calibrate the collagen clock.

#### *Comparative biochemistry and the origin and early evolution of the Metazoa*

A different approach to the problem of the early history of the Metazoa has been explored by Towe (1970, 1981). He has attempted to use the distribution of certain molecules in the living biota to determine relationships between different distantly related groups of organisms. He suggested, for example, that as collagen is limited to the Metazoa and requires molecular oxygen for its production (Kikuchi *et al.* 1983), the time of origin of fossilizable animals was determined by oxygen levels in the atmosphere and hydrosphere (Towe 1970, 1981; Runnegar 1982*b*, *c*).

The molecular evidence for historical relationships between the animal phyla has not been explored in any comprehensive way. In part, this is because there is too little information available, but it is also due to the fact that no systematic study of the available data has yet been made. However, it is possible to suggest some methods of approach and to identify some of the potentially useful molecules.

Most genes contain many characters in addition to their primary structures (DNA sequences). These characters include the position and nature of promoter sequences; the presence/absence and size of signal and/or propeptides; the sizes and positions of exons, protein-coding regions, and introns; the presence of tandem repeats or palindromes; and the position and nature of polyadenylation signals. Similarly, each of the proteins specified by homologous genes may display differences in their secondary or higher order structures, active-site ligands, hydrophobic regions, etc. Each character is therefore of potential phylogenetic significance and may be analysed in a cladistic fashion.

For example, at the time of his death, Tom Schopf was attempting to use the positions of the non-coding sequences (introns) in the actin genes of eukaryotes to examine relationships between the animal phyla (pers. comm. 10 February 1984). In vertebrate  $\alpha$  and  $\beta$  actin genes the introns lie within or adjacent to codons 41, 121, 150, 204, 267, and 327 (Nudel *et al.* 1983) and a similar arrangement is found in sea urchin actin genes (codons 41, 121, 204, and 267). By contrast, arthropod and nematode actin genes have introns in different positions. Thus, this evidence supports the close relationship of



echinoderms and vertebrates and suggests that uniramian arthropods, echinoderms/vertebrates, and nematodes are equally distant.

Because small animals can respire by simple diffusion (Alexander 1971) they do not require respiratory transport pigments. The evolution of these complexes must therefore post-date the origin of the Metazoa and the evolution of collagen (which was needed to build bigger bodies). An understanding of the evolutionary histories of the respiratory transport pigments should therefore provide some insight into the early history of the Metazoa. For example, hemerythrin is an intracellular non-haeme oxygen carrier that has Fe at the active site. It has been found in sipunculids, priapulids, *Lingula*, and one annelid, but is only well known from the Sipunculida (Klotz *et al.* 1976). A determination of the amino-acid sequence of lingulid hemerythrin (Joshi and Sullivan 1973) should therefore provide important information about the relationship of the two phyla and their affinities with the proterostomes and deuterostomes. Similarly, it would be a big help to have the amino-acid/gene sequences of platyhelminth, nemertean, nematode and holothurian globins, and molluscan haemocyanins. This kind of information should become available within the next decade.

### BIOMINERALIZATION AND MOLECULAR BIOLOGY

Many different kinds of organisms deposit crystalline or amorphous inorganic compounds inside or outside their cells (Lowenstam 1981). These biominerals are frequently used to construct rigid skeletons, but they are also used to strengthen flexible walls, to rid the cells of unwanted salts, to store useful ions, or to form parts of sensory organs used for sight, orientation, and navigation. As the formation and organization of biominerals occurs primarily at the molecular level, studies of the production, construction, and preservation of mineral skeletons and other biominerals link molecular biology with palaeontology and other branches of geology. For example, Riding (1982) has suggested that the late Proterozoic–Jurassic fossil record of calcified marine cyanophytes may reflect lower Mg/Ca ratios in sea water during that period (but see Sandberg 1983), and Cook and Shergold (1984) have argued that the time of origin and composition of the skeletons of Cambrian invertebrates are related to major changes in the concentration of phosphate in the shallower parts of the early Cambrian oceans.

If these kinds of useful hypotheses are to be generated and tested, it will be necessary to understand much more about the mechanisms and history of biomineralization. The generalization that phosphate skeletons were common in the Cambrian and rare thereafter (Lowenstam and Margulis 1980) needs to be explored further through petrographic, SEM, and electron microprobe studies of Cambrian fossils. In addition the recent discovery that original skeletal microstructures are frequently replicated by phosphatic internal moulds (Runnegar and Bentley 1983; Runnegar 1983, *in press b*) should make it possible to determine the nature and composition of carbonate skeletons in which the original microstructures have been destroyed by recrystallization. There is already good evidence that other fine-grained casting media (e.g. dolomitized micrite) may also yield excellent replicas of original microstructures (J. Pojeta, Jr., pers. comm.).

It is possible to gain some insight into the molecular controls on skeletal construction by examining skeletons formed from well-ordered crystalline subunits. If the crystalline subunits have a form or an arrangement which is not found in natural crystals of the same mineral, it is fairly easy to identify the effect—and perhaps the cause—of the biological control.

For example, natural inorganic crystals of calcite are known to develop some 328 different crystallographic forms (Runnegar 1984c). However, most natural and synthetic crystals and many biominerals display only the most common forms, normally low-index rhombohedra, and simple prisms. In contrast the mica-like calcite folia of the window-pane shell *Placuna placenta* have their surfaces constructed from the very rare rhombohedral form  $\{10\bar{1}8\}$  (Runnegar 1984c). This form lies perpendicular to a direction of fast crystal growth and should not appear under normal conditions. Its extreme development in *P. placenta* is therefore clearly under biological control, and probably results from a two-dimensional stereo-chemical similarity between the mineral lattice and the

interleaved protein matrix; the array of calcium atoms in the  $\{10\bar{1}8\}$  plane of the calcite lattice matches the inter-residue dimensions of a parallel  $\beta$ -pleated sheet of a protein with a repetitive amino-acid sequence of the form (glycine-aspartic acid)<sub>n</sub> (Runnegar 1984c). Thus, the crystallography of the mineral phase may, at least in principle, be used to determine something about the nature of the adjacent organic matrix. It should, therefore, be possible to make the same kinds of deductions from the skeletons of extinct organisms.

The organization of individual skeletal elements may also be informative. The coccosphere of the 'living fossil' *Braarudosphaera bigelowi* is a regular dodecahedron formed of twelve equal-sized pentaliths. Each pentalith is composed of five calcite crystals arranged in a highly organized way (Runnegar, in press c). As pentagonal symmetry does not exist in the calcite lattice the assembly of the pentaliths in the Golgi apparatus of the algal cells must be specified at the molecular level. If the operation of such systems were well understood, it would again be possible to make some precise deductions from the skeletons of extinct organisms.

Stefan Bengtson (pers. comm.) has pointed out that it would be useful to be able to distinguish between collagen-mediated phosphate skeletons and those formed in other ways. So far as is known, vertebrate bone and teeth are mineralized by the deposition of apatite within and between collagen fibrils (Lees 1979; Höhling *et al.* 1980; Glimcher 1984). On the other hand, phosphate deposition in the muscles of the polychaete *Nephtys* (Gibbs and Bryan 1984) and the periostracum of the bivalve *Lithophaga* (Waller 1983) obviously occurs in a fundamentally different manner. Are these various modes of phosphatization distinguishable microscopically? Could we, for example, identify a mineralized analogue of the graptolite periderm? These are the kinds of questions that need to be answered if we are to understand the true nature of the phosphatic microfossils found in early Cambrian strata.

#### *Determination of genome sizes*

Despite the difficulties of measurement there is considerable evidence that the genomic DNA content (haploid content, C-value) of organisms is correlated with cell volume and nuclear volume (Cavalier-Smith 1978). Thomson (1972) used this relationship to show that the abnormally high (diploid) DNA contents of living lungfish (160–285 picograms per cell; Pedersen 1971) were developed slowly throughout the evolutionary history of the lungfish. He based his analysis on measurements of the dimensions of osteocyte lacunae in fossil bone. The osteocytes of Devonian lungfish were found to be less than a tenth of the volume of the osteocytes of living lepidosirenid lungfish, and their DNA content is therefore likely to have been comparable to that found in living mammals (3–5 pg per cell). As most of the morphological innovations occurred early in the history of the group (Campbell and Barwick 1983) the great increase in DNA content followed rather than caused the rapid evolution of this group of organisms (Thomson 1972).

It would be useful to have more data of this kind to test the generalization that exceptionally large amounts of DNA are found in the genomes of 'living fossils' (Hinegardner 1976). Such a correlation may imply that the extra 'junk' DNA somehow stifles evolutionary change (Thomson 1972) although other explanations are also possible (Grime and Mowforth 1982).

The obvious problem for palaeontologists is the determination of cell size, but there is also a need for more information from living organisms and for a more precise method of comparing estimates of DNA content (Greilhuber *et al.* 1983). Measuring cell size is not difficult in permineralized plants, but it is not easy to estimate the cell sizes of extinct animals. Nevertheless, there may be ways to tackle this problem. The studies of Pawlicki (1984a, b) on dinosaur bones show that osteocytes may be spectacularly preserved; the fact that each shell prism of the secondary shell layer of living rhynchonellid and terebratulid brachiopods is formed by a single epithelial cell (Williams 1968) may enable epithelial cell size to be determined in fossil brachiopods; and the report (Giraud-Guille 1984) that the outlines of the epidermal cells of the crab are reflected in structures that penetrate the cuticle may indicate that epidermal cell size can be measured in extinct arthropods. It remains to be seen whether these and other ways of measuring cell size in fossil invertebrates will prove to be practical.

## CONCLUDING REMARKS

The idea that genome size may be reflected in the dimensions of the calcite prisms of the shells of brachiopods may seem far-fetched but it leads to the question of how far up the morphological hierarchy should we expect molecular structures to persist? I have already shown that the basic organization of collagen genes is reflected in the cross-banding of collagen fibrils and that the geometry of matrix proteins may control the most fundamental property of the shell of *P. placenta*. In *B. bigleowi* the topology of a particular macromolecule (or set of macromolecules) appears to be responsible for the design of the whole exoskeleton (Runnegar, in press *c*) and the same may be true for most coccolith-bearing algae. Viewed in this way, there is no fundamental difference between molecular biology and classical anatomy and it is important to try to integrate the knowledge of both disciplines. Some of the recent results of developmental biology (Davidson *et al.* 1982; Sánchez-Herrero *et al.* 1985; Fjose *et al.* 1985) are providing the first steps in this direction.

There is also a need to bridge the gulf between molecular biology and classical population genetics (Doolittle 1982) and a need for new general statements about the way evolution has occurred. We know a great deal about the processes that lead to new species but rather little about the processes that give rise to fundamentally new structures or new kinds of molecules (Jaanusson 1981; Runnegar 1984*b*). It is not clear whether rapid rates or significant amounts of morphological change are accompanied by comparable changes in the information content of the genome or whether the information is merely rearranged in some way. Commenting on a similar point, Doolittle (1982, p. 88) wrote: 'We must disabuse ourselves of the notion that organisms considered "primitive" because of their morphological and behavioural simplicity have primitive molecular biologies. Just the opposite may well be true.' And finally we know little of the role of gene transfer between unrelated lineages in the development and diversification of life.

As one molecular biologist said to me recently, cytochrome *c* is a boring molecule; it has changed little in billions of years (Dickerson 1980). By this he meant that the morphology of the molecule is highly conserved; there are considerable differences in the messages of the genes encoding the protein in different lineages. This apparent paradox is easily explained by the fact that there are so many possible solutions to the same problem. There are about 100 amino-acid residues in an average-sized molecule of cytochrome *c*, and only about three of these residues are fully conserved in the proteins so far studied. If only one to two closely related amino acids could occupy each of the remaining sites, there would still be about  $2^{100}$  possible combinations. It is, therefore, obvious that there is enormous scope for genomic evolution with little or no effect upon morphology. Thus, it is more important to distinguish between the evolution of information and the evolution of molecular and anatomical structures rather than to attempt to isolate 'molecular evolution' from phenomena observed at higher morphological levels.

The current dogma of molecular evolution is essentially gradualistic; a succession of small changes in the amino-acid sequences gradually converts one kind of protein into another. This mechanism explains the evolution of closely related proteins but it does not adequately account for the origin of fundamentally new kinds of enzymes. These may well arise by the fusion of parts of two or more unrelated genes (Guiard and Lederer 1979; Runnegar 1984*b*); the new product may have a new morphology and a new function and yet be specified by old information.

A striking example of the non-gradualistic evolution of a new enzyme is given by Ohno (1984). Despite the fact that the industrial synthesis of nylon began only several decades ago, it was found in 1975 that a species of *Flavobacterium* could grow in a culture medium containing a by-product of nylon factories (6-aminohexanoic acid cyclic dimer) as the sole source of carbon and nitrogen. Ohno has suggested that the new enzyme arose by the insertion of a single nucleotide into the beginning of a pre-existing protein-coding sequence. This insertion resulted in a change in the DNA reading frame allowing the old information to be translated in a fundamentally new way.

Compared with the number of living and extinct species the number of extant and extinct enzymes and other kinds of proteins is relatively small and many are widely shared amongst distant taxa. It is, therefore, clear that the evolution of an important new kind of protein (e.g. collagen) has always been



a rare event. Given this fact, how likely is it that genes for useful proteins have been acquired by lateral transfer from unrelated organisms?

The best possible example of this phenomenon is the presence of haemoglobin genes in leguminous and non-leguminous angiosperms (Brisson and Verma 1982; Appleby *et al.* 1983; Kortt *et al.* 1985), but, so far as is known, in no other plants. There is ample evidence that the plant haemoglobins are homologous to animal globins (Runnegar 1984*b*) and so there are two alternatives: either the angiosperms inherited the gene from the common ancestor of animals and plants and it remained unexpressed for hundreds of millions of years or a haemoglobin gene was transferred from an animal to an angiosperm some tens of millions of years ago.

The plant haemoglobins are restricted to nitrogen-fixing root nodules formed in a symbiotic association with the bacterium *Rhizobium*. The haeme moiety appears to be manufactured by the bacterium whereas the protein is synthesized by the plant after bacterial infection has occurred (Ellfolk 1972; Dilworth and Glenn 1984). The haemoglobin occurs in the cytoplasm and nucleus of the infected cells but not in the bacteroids or the peribacteroid spaces (Robertson *et al.* 1984).

Given this highly specific and unusual association it is tempting to conclude that the angiosperms have acquired the haemoglobin gene by lateral transfer. The difficulty with this interpretation results from the fact that the amino-acid sequences of the angiosperm globins are about 80–90% different from all animal globins so far sequenced. Thus, unless the angiosperm globins have been evolving at a much faster rate than animal globins, the sequence differences indicate that the animal and plant globins diverged about a billion years ago.

A faster rate of evolution is a distinct possibility given the substantial differences (up to 60%) in the amino-acid sequences of different angiosperm globins (Kortt *et al.* 1985). However, another possibility is that the ancestral plant globin gene was obtained from a member of an invertebrate phylum such as the Nematoda. Until at least one globin sequence is available from a representative of each of the phyla that could have contributed a globin gene to the angiosperms, it will be difficult to exclude the possibility that the angiosperm genes were obtained through lateral transfer.

If such lateral gene transfers have occurred during the course of evolution, they may well represent very rare events. However, given the length of geological time, even very rare events have a high probability of occurrence. It is, therefore, important for students of evolution to consider the implications of such rare events in the evolution and development of life. It may no longer be possible to assume *a priori* that demonstrably homologous characters are necessarily confined to single clades.

*Acknowledgements.* I thank Maria Runnegar for much encouragement, advice, and help with the literature of biochemistry. The computing and other technical assistance was provided by F. A. Shaw and the manuscript typed by R. K. Vivian. Much of this work was supported financially by the Australian Research Grants Scheme.

## REFERENCES

- ABELSON, P. H. 1956. Paleobiochemistry. *Sci. Am.* **195** (1), 2–7.
- ADAMS, E. 1978. Invertebrate collagens. *Science*, **202**, 591–598.
- ALEXANDER, R. M. 1971. *Size and shape*, 59 pp. Edward Arnold, London.
- ANDERSON, D. T. 1982. Origins and relationships among the animal phyla. *Proc. Linn. Soc. N.S.W.* **106**, 151–166.
- APPLEBY, C. A., TJEPKEMA, J. D. and TRINICK, M. J. 1983. Hemoglobin in a nonleguminous plant, *Parasponia*: possible genetic origin and function in nitrogen fixation. *Science*, **220**, 951–953.
- ARMSTRONG, W. G., DILLY, P. N. and URBANEK, A. 1984. Collagen in the pterobranch coenecium and the problem of graptolite affinities. *Lethaia*, **17**, 145–152.
- HALSTEAD, L. B., REED, F. B. and WOOD, L. 1983. Fossil proteins in vertebrate calcified tissues. *Phil. Trans. R. Soc. Lond.* **B301**, 301–343.
- ASHFORD, D. and NEUBERGER, A. 1980. 4-hydroxyl-L-proline in plant glycoproteins. *Trends biochem. Sci.* **5**, 245–248.
- AUDLEY-CHARLES, M. G. 1983. Reconstruction of eastern Gondwanaland. *Nature*, **306**, 48–50.
- AYALA, F. J. 1978. The mechanisms of evolution. *Sci. Am.* **293** (3), 56–69.
- BENTZ, H., BÄCHINGER, H. P., GLANVILLE, R. and KÜHN, K. 1978. Physical evidence for the assembly of A and B chains of human placental collagen in a single triple helix. *Eur. J. Biochem.* **92**, 563–567.

- BERNARD, M. P., CHU, M.-L., MYERS, J. C., RAMIREZ, F., EIKENBERRY, E. F. and PROCKOP, J. 1983. Nucleotide sequences of complementary deoxyribonucleic acids for the Pro $\alpha$ 1 chain of human type I procollagen. Statistical evaluation of structures that are conserved during evolution. *Biochemistry*, **22**, 5213–5223.
- BERTHET-COLOMINAS, C., MILLER, A. and WHITE, S. W. 1979. Structural study of the calcifying collagen in turkey leg tendons. *J. molec. Biol.* **134**, 431–445.
- BODENMÜLLER, H. and SCHALLER, H. C. 1981. Conserved amino-acid sequence of a neuropeptide, the head activator, from coelenterates to humans. *Nature*, **293**, 579–580.
- BORNSTEIN, P. and SAGE, H. 1980. Structurally distinct collagen types. *Ann. Rev. Biochem.* **49**, 957–1003.
- BRIGGS, D. E. G. and FORTEY, R. A. 1982. The cuticle of the aglaspidid arthropods, a red-herring in the early history of the vertebrates. *Lethaia*, **15**, 25–29.
- BROOKS, J. (ed.) 1981. *Organic maturation studies and fossil fuel exploration*, 441 pp. Academic Press, London.
- BROWN, R. D. 1980. The origin of life on earth. In PEACOCK, W. J. (ed.). *Life: hierarchies of interacting molecules*, 1–6. Silver Jubilee Symposium Volume 111. Australian Academy of Science, Canberra.
- CALVIN, M. 1968. Molecular palaeontology. *Trans. Leicester Literary Phil. Soc.* **62**, 45–69.
- CAMPBELL, K. S. W. and BARWICK, R. E. 1983. Early evolution of dipnoan dentitions and a new genus *Sponesyrion*. *Mem. Ass. Australas. Palaeontols*, **1**, 17–49.
- CAVALIER-SMITH, T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J. cell Sci.* **34**, 247–278.
- COOK, P. J. and SHERGOLD, J. H. 1984. Phosphorus, phosphorites and skeletal evolution at the Precambrian–Cambrian boundary. *Nature*, **308**, 231–236.
- CROWTHER, P. and RICKARDS, B. 1977. Cortical bandages and the graptolite zooid. *Geologica Palaeontol.* **11**, 9–46.
- DAVIDSON, E. H., HOUGH-EVANS, B. R. and BRITTEN, R. J. 1982. Molecular biology of the sea urchin embryo. *Science*, **217**, 17–26.
- DEGENS, E. T. 1967. Evolutionary trends inferred from the organic tissue variation of mollusc shells. *Medd. Fra Dansk Geol. Forening*, **17**, 112–124.
- DEVORE, D. P., ENGBRETSON, G. H., SCHACHTELE, C. F. and SAUK, J. J. 1984. Identification of collagen from byssus threads produced by the sea mussel, *Mytilus edulis*. *Comp. Biochem. Physiol.* **77B**, 529–531.
- DICKERSON, R. E. 1980. Cytochrome *c* and the evolution of energy metabolism. *Sci. Am.* **242** (3), 136–153.
- and GEIS, I. 1983. Hemoglobin: structure, function, evolution and pathology, 176 pp. Benjamin/Cummings Publishing, Menlo Park, Calif.
- DILWORTH, M. and GLENN, A. 1984. How does a legume nodule work? *Trends biochem. Sci.* **9**, 519–523.
- DOOLITTLE, W. F. 1982. Evolutionary molecular biology: where is it going? *Can. J. Biochem.* **60**, 83–90.
- DOSE, K. 1983. Prebiotic polypeptides and the origin of biological information. *Naturwiss.* **70**, 378–382.
- EGLINTON, G. and CALVIN, M. 1967. Chemical fossils. *Sci. Am.* **216** (1), 32–43.
- EKSTROM, A., FOKES, C. J. R., HAMBLEY, T., LOEH, H. J., MILLER, S. A. and TAYLOR, J. C. 1983. Determination of the crystal structure of a petroporphyrin isolated from oil shale. *Nature*, **306**, 173–174.
- ELLFOLK, N. 1972. Leghaemoglobin, a plant haemoglobin. *Endeavour*, **31**, 139–142.
- FJOSE, A., MCGINNIS, W. J. and GEHRING, W. J. 1985. Isolation of a homoeo box-containing gene from the engrailed region of *Drosophila* and the spatial distribution of its transcripts. *Nature*, **313**, 284–289.
- FRASER, R. D. B., MACRAE, T. P. and SUZUKI, E. 1979. Chain conformation in the collagen molecule. *J. molec. Biol.* **129**, 463–481.
- FRAUENFELDER, H. 1983. Summary and outlook. Mobility and function in proteins and nucleic acids. *Ciba Foundation Symp.* **93**, 329–343.
- GALLOWAY, J. 1984. Are biological structures piles of molecules or just heaps? *Trends. biochem. Sci.* **9**, 233–238.
- GIBBS, P. E. and BRYAN, G. W. 1984. Calcium phosphate granules in muscle cells of *Nephtys* (Annelida, Polychaeta)—a novel skeleton? *Nature*, **310**, 494–495.
- GINGERICH, P. D. 1984. Fossils, molecular clocks, and origin of animal phyla. *97th Annual Meeting Geol. Soc. Am. Abstracts with Programs*, **16**, 519.
- GIRAUD-GUILLE, M. M. 1984. Calcification initiation sites in the crab cuticle: the interprismatic septa. *Cell Tissue Res.* **236**, 413–420.
- GLIMCHER, M. J. 1984. Recent studies of the mineral phase in bone and its possible linkage to the organic matrix by protein-bound phosphate bonds. *Phil. Trans. R. Soc. Lond.* **B304**, 479–508.
- GOLDING, G. B. 1983. Estimates of DNA and protein sequence divergence: an examination of some assumptions. *Mol. Biol. Evol.* **1**, 125–142.
- GOODMAN, M. 1981. Globin evolution was apparently very rapid in early vertebrates: a reasonable case against the rate-constancy hypothesis. *J. molec. Evol.* **17**, 114–120.

- GOULD, R. E. 1971. *Lysoxylon grigsbyi*, a cycad trunk from the Upper Triassic of Arizona and New Mexico. *Amer. J. Bot.* **58**, 239–248.
- GOULD, S. J. 1984. The life and work of T. J. M. Schopf (1939–1984). *Paleobiology*, **10**, 280–285.
- GREILHUBER, J., VOLLETH, M. and LOIDL, J. 1983. Genome size of man and animals relative to the plant *Allium cepa*. *Can. J. Genet. Cytol.* **25**, 554–560.
- GRIME, J. P. and MOWFORTH, M. A. 1982. Variation in genome size—an ecological interpretation. *Nature*, **299**, 151–153.
- GUIARD, B. and LEDERER, F. 1979. The 'cytochrome b<sub>5</sub> fold': structure of a novel protein superfamily. *J. molec. Biol.* **135**, 639–650.
- HADLER, H. I., DIMITRIJEVIC, B. and MAHALINGAM, R. 1983. Mitochondrial DNA and nuclear DNA from normal rat liver have a common sequence. *Proc. Natl Acad. Sci. USA*, **80**, 6495–6499.
- HENKEL, W. and GLANVILLE, R. W. 1982. Covalent crosslinking between molecules of type I and type III collagen. *European J. Biochem.* **122**, 205–213.
- HIGUCHI, R., BOWMAN, B., FREIBERGER, M., RYDER, O. A. and WILSON, A. C. 1984. DNA sequences from the quagga, an extinct member of the horse family. *Nature*, **312**, 282–284.
- and WILSON, A. C. 1984. Recovery of DNA from extinct species. *Federation Proceedings*, **43**, 1557.
- HINEGARDNER, R. 1976. Evolution of genome size. In AYALA, F. J. (ed.). *Molecular evolution*, 179–199. Sinauer Associates, Sunderland, Massachusetts.
- HÖHLING, H. J., BARCKHAUS, R. H. and KREFTING, E. R. 1980. Hard tissue formation in collagen-rich systems: calcium phosphate nucleation and organic matrix *Trends biochem. Sci.* **5**, 8–11.
- JAANUSSON, V. 1981. Functional thresholds in evolutionary progress. *Lethaia*, **14**, 251–260.
- JOSHI, J. G. and SULLIVAN, B. 1973. Isolation and preliminary characterization of hemerythrin from *Lingula unguis*. *Comp. Biochem. Physiol.* **44B**, 857–867.
- KIKUCHI, Y., SUZUKI, Y. and TAMIYA, N. 1983. The source of oxygen in the reaction catalysed by collagen lysyl hydroxylase. *Biochem. J.* **213**, 507–512.
- KIMBER, R. W. L. and MILNES, A. R. 1984. The extent of racemization of amino acids in Holocene and Pleistocene marine molluscs in southern South Australia: preliminary data on a time-framework for calcrete formation. *Aust. J. Earth Sci.* **31**, 279–286.
- KING, K. 1978.  $\gamma$ -carboxyglutamic acid in fossil bones and its significance for amino acid dating. *Nature*, **273**, 41–43.
- KLOTZ, I. M., KLIPPENSTEIN, G. L. and HENDRICKSON, W. A. 1976. Hemerythrin: alternative oxygen carrier. *Science*, **192**, 335–344.
- KORTT, A. A., BURNS, J. E., TRINICK, M. J. and APPLEBY, C. A. 1985. The amino acid sequence of hemoglobin I from *Parasponia andersonii*, a nonleguminous plant. *FEBS Lett.* **180**, 55–60.
- LAWRENCE, J. R. and WHITE, J. W. C. 1984. Growing season precipitation from D/H ratios of Eastern White Pine. *Nature*, **311**, 558–560.
- LEES, S. 1979. A model for the distribution of HAP crystallites in bone—an hypothesis. *Calcif. Tiss. Intl*, **27**, 53–56.
- LEONARDI, L., RUGGERI, A., ROVERI, N., BIGI, A. and REALE, E. 1983. Light microscopy, electron microscopy, and X-ray diffraction analysis of glycerinated collagen fibres. *J. ultrastruct. Res.* **85**, 228–237.
- LESK, A. M. and CHOTHIA, C. 1980. How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. molec. Biol.* **136**, 225–270.
- LODISH, H. F. and ROTHMAN, J. E. 1979. The assembly of cell membranes. *Sci. Am.* **240** (1), 38–53.
- LOWENSTAM, H. A. 1981. Minerals formed by organisms. *Science*, **211**, 1126–1131.
- and MARGULIS, L. 1980. Evolutionary prerequisites for early Phanerozoic calcareous skeletons. *BioSystems*, **12**, 27–41.
- LOWENSTEIN, J. M. 1980. Species-specific proteins in fossils. *Naturwiss.* **67**, 343–346.
- MOLLESON, T. and WASHBURN, S. L. 1982. Piltdown jaw confirmed as orang. *Nature*, **299**, 294.
- and ZIHLMAN, A. L. 1984. Human evolution and molecular biology. *Perspectives Biol. Med.* **27**, 611–622.
- MACKENZIE, A. S., BRASSELL, S. C., EGLINTON, G. and MAXWELL, J. R. 1982. Chemical fossils: the geological fate of steroids. *Science*, **217**, 491–504.
- MCLACHLAN, A. D. 1976. Evidence for gene duplication in collagen. *J. molec. Biol.* **107**, 159–174.
- MATSUNO, K. 1983. Evolutionary changes in the information content of polypeptides. *J. theor. Biol.* **105**, 185–199.
- MATTAJ, I. W. 1984. snRNAs: from gene architecture to RNA processing. *Trends biochem. Sci.* **9**, 435–437.



- MATTHEWS, B. W., FENNA, R. E., BOLOGNESI, M. C., SCHMID, M. F. and OLSON, J. M. 1979. Structure of a bacteriochlorophyll *a*-protein from the green photosynthetic bacterium *Prosthecochloris aestuarii*. *J. molec. Biol.* **131**, 259–285.
- MESSERSCHMIDT, U., WILHELM, P., PILZ, I., KAPP, O. H. and VINOGRADOV, S. N. 1983. The molecular size and shape of the extracellular hemoglobin of *Nephtys incisa*. *Biochim. Biophys. Acta*, **742**, 366–373.
- MONSON, J. M., NATZLE, J., FRIEDMAN, J. and MCCARTHY, B. J. 1982. Expression and novel structure of a collagen gene in *Drosophila*. *Proc. Natl Acad. Sci. USA*, **79**, 1761–1765.
- NIKLAS, K. J., TIFFNEY, B. H. and LEOPOLD, A. C. 1982. Preservation of unsaturated fatty acids in Palaeogene angiosperm fruits and seeds. *Nature*, **296**, 63–64.
- NORDWIG, A. and HAYDUK, U. 1969. Invertebrate collagens: isolation, characterization and phylogenetic aspects. *J. molec. Biol.* **44**, 161–172.
- NUDEL, U., ZAKUT, R., SHANI, M., NEUMAN, S., LEVY, Z. and YAFFE, D. 1983. The nucleotide sequence of the rat cytoplasmic  $\beta$ -actin gene. *Nucleic Acids Res.* **11**, 1759–1771.
- OHNO, S. 1984. Birth of a unique enzyme from an alternative reading frame of the preexisted, internally repetitive coding sequence. *Proc. Natl Acad. Sci. USA*, **81**, 2421–2425.
- OURISSON, G., ALBRECHT, P. and ROHMER, M. 1979. The hopanoids. Palaeochemistry and biochemistry of a group of natural products. *Pure Appl. Chem.* **51**, 709–729.
- 1982. Predictive microbial biochemistry— from molecular fossils to procaryotic membranes. *Trends biochem. Sci.* **7**, 236–239.
- 1984. The microbial origin of fossil fuels. *Sci. Am.* **251** (2), 34–41.
- PAWLICKI, R. 1984a. Metabolic pathways of the fossil dinosaur bones. Part III. Intermediary and other osteocytes in the system of metabolic pathways of dinosaur bone. *Folia Histochem. Cytobiol.* **22**, 91–98.
- 1984b. Metabolic pathways of the fossil dinosaur bones. Part IV. Modes of linkage between osteocytes and a variety of nexuses of osteocytes processes. *Ibid.* 99–104.
- PEDERSEN, R. A. 1971. DNA content, ribosomal gene multiplicity, and cell size in fish. *J. exp. Zool.* **177**, 65–78.
- PHILP, R. P. and GILBERT, T. D. 1982. Unusual distribution of biological markers in an Australian crude oil. *Nature*, **299**, 245–247.
- PROUDFOOT, N. J., GIL, A. and MANIATIS, T. 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical pseudogene. *Cell*, **31**, 553–563.
- RICHARDSON, J. S. 1981. The anatomy and taxonomy of protein structure. *Adv. protein Chem.* **34**, 167–330.
- RIDING, R. 1982. Cyanophyte calcification and changes in ocean chemistry. *Nature*, **299**, 814–815.
- ROBERTSON, J. G., WELLS, B., BISSELING, T., FARNDEN, K. J. F. and JOHNSTON, A. W. B. 1984. Immuno-gold localization of leghaemoglobin in cytoplasm in nitrogen-fixing root nodules of pea. *Ibid.* **311**, 254–256.
- RUNNEGAR, B. 1982a. A molecular-clock date for the origin of the animal phyla. *Lethaia*, **15**, 199–205.
- 1982b. Oxygen requirements, biology and phylogenetic significance of the late Precambrian worm *Dickinsonia*, and the evolution of the burrowing habit. *Alcheringa*, **6**, 223–239.
- 1982c. The Cambrian explosion: animals or fossils? *J. geol. Soc. Aust.* **29**, 395–411.
- 1983. Molluscan phylogeny revisited. *Mem. Ass. Australas. Palaeontols*, **1**, 121–144.
- 1984a. The Permian of Gondwanaland. *Proc. 27th Int. Geol. Cong.* **1**, 305–339.
- 1984b. Derivation of the globins from type *b* cytochromes. *J. molec. Evol.* **21**, 33–41.
- 1984c. Crystallography of the foliated calcite shell layers of bivalve molluscs. *Alcheringa*, **8**, 273–290.
- in press *a*. Collagen gene construction and evolution. *J. molec. Evol.*
- in press *b*. Shell microstructures of Cambrian molluscs replicated by phosphate. *Alcheringa*.
- in press *c*. Crystallography and biomineralization of the dodecahedral coccosphere of *Braarudosphaera bigelowi* (Gran and Braarud). *Palaeontology*.
- and BENTLEY, C. 1983. Anatomy, ecology and affinities of the Australian Early Cambrian bivalve *Pojetaia runnegari* Jell. *J. Paleont.* **57**, 73–92.
- SÁNCHEZ-HERRERO, E., VERNÓS, I., MARCO, R. and MORATA, G. 1985. Genetic organization of *Drosophila* bithorax complex. *Nature*, **313**, 108–113.
- SANDBERG, P. A. 1983. An oscillating trend in Phanerozoic non-skeletal carbonate mineralogy. *Ibid.* **305**, 19–22.
- SAXBY, J. D. 1983. The distribution, properties and potential of Toolebuc oil shale. *Proceedings of the first Australian workshop on oil shale*, 43–46. Lucas Heights.
- SCHOPF, J. W. 1978. The evolution of the earliest cells. *Sci. Am.* **239** (3), 110–138.
- SEPKOSKI, J. J. 1978. A kinetic model of Phanerozoic taxonomic diversity. I. Analysis of marine orders. *Paleobiology*, **4**, 223–251.

- SMITH, J. W., RIGBY, D., SCHMIDT, P. W. and CLARK, D. A. 1983. D/H ratios of coals and the palaeolatitude of their deposition. *Nature*, **302**, 322–323.
- STERN, D. B. and PALMER, J. D. 1984. Extensive and widespread homologies between mitochondrial DNA and chloroplast DNA in plants. *Proc. Natl Acad. Sci. USA*, **81**, 1946–1950.
- STORM, C. B., KRANE, J., SKJETNE, T., TELNAES, N., BRANTHAVER, J. F. and BAKER, E. W. 1984. The structure of abelsonite. *Science*, **223**, 1075–1076.
- SYLVESTER-BRADLEY, P. 1964. The origin of oil—and life. *Discovery* **25** (5), 37–42.
- TATE, V., FINER, M., BOEDTKER, H. and DOTY, P. 1983. Procollagen genes: further sequence studies and interspecies comparisons. *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1039–1049.
- THOMSON, K. S. 1972. An attempt to reconstruct evolutionary changes in the cellular DNA content of lungfish. *J. exp. Zool.* **180**, 363–372.
- THORNBER, J. B. and MARKWELL, J. P. 1981. Photosynthetic pigment-protein complexes in plant and bacterial membranes. *Trends biochem. Sci.* **6**, 122–125.
- TOWE, K. M. 1970. Oxygen-collagen priority and the early metazoan fossil record. *Proc. Natl Acad. Sci. USA*, **65**, 781–788.
- 1981. Biochemical keys to the emergence of complex life. In BILLINGHAM, J. (ed.). *Life in the Universe*, 297–306. MIT Press, Cambridge, Massachusetts.
- and URBANEK, A. 1972. Collagen-like structures in Ordovician graptolite periderm. *Nature*, **237**, 443–445.
- TRAUB, W. and YONATH, A. 1966. Polymers of tripeptides as collagen models. I. X-ray studies of poly (L-prolyl-glycyl-L-proline) and related polytripeptides. *J. molec. Biol.* **16**, 404–414.
- URBANEK, A. and TOWE, K. M. 1974. Ultrastructural studies on graptolites, 1: the periderm and its derivatives in the Dendroidea and in *Mastigograptus*. *Smithsonian Contrib. Paleobiol.* **20**, 1–48.
- 1975. Ultrastructural studies on graptolites, 2: the periderm and its derivatives in the Graptoloidea. *Ibid.* **22**, 1–24.
- WALLER, T. R. 1983. Dahllite in the periostracum of *Lithophaga nigra* (Mollusca: Bivalvia) and its taxonomic and functional implications. *Amer. malacol. Bull.* **1**, 101.
- WALTER, M. R. and HEYS, G. R. in press. Links between the rise of the Metazoa and the decline of stromatolites. *Precambrian Res.*
- WALTON, A. G. 1981. *Polypeptides and protein structure*. 393 pp. Elsevier, New York.
- WILLIAMS, A. 1968. *Evolution of the shell structure of articulate brachiopods*. *Spec. Pap. Palaeontol.* **2**, 1–55.
- WILSON, A. C., CARLSON, S. S. and WHITE, T. J. 1977. Biochemical evolution. *Ann. Rev. Biochem.* **46**, 573–639.
- WOODHEAD-GALLOWAY, J. 1980. *Collagen: the anatomy of a protein*, 59 pp. Edward Arnold, London.
- YAMADA, Y., AVVEDIMENTO, V. E., MUDRYJ, M., OHKUBO, H., VOGELI, G., IRANI, M., PASTAN, I. and DECROMBRUGGHE, B. 1980. The collagen gene: evidence for its evolutionary assembly by amplification of a DNA segment containing an exon of 54 bp. *Cell*, **22**, 887–892.
- ZUCKERKANDL, E. and PAULING, L. 1962. Molecular disease, evolution and genetic heterogeneity, In KASHA, M. and PULLMAN, B. (eds.). *Horizons in biochemistry*, 189–225. Academic Press, New York.

BRUCE RUNNEGAR

Department of Geology and Geophysics  
University of New England  
Armidale, New South Wales  
Australia, 2351

Typescript received 15 March 1985

Revised typescript received 21 June 1985